



Institute for Empirical Research in Economics
University of Zurich

Working Paper Series
ISSN 1424-0459

Working Paper No. 333

**Harmonic Regression Models: A Comparative
Review with Applications**

Michael Artis, José G. Clavel, Mathias Hoffmann and Dilip
Nachane

September 2007

Harmonic Regression Models: A Comparative Review with Applications.

Michael Artis

Manchester Regional Economics Centre, University of Manchester, UK.

José G. Clavel*

University of Murcia, Spain.

Mathias Hoffmann

University of Zurich, Zurich, Switzerland.

Dilip Nachane

Indira Gandhi Institute of Development Research, India.

September 26, 2007

Abstract

Strongly periodic series occur frequently in many disciplines. This paper reviews one specific approach to analyzing such series viz. the *harmonic regression approach*. In this paper the five major methods suggested under this approach are critically reviewed and compared, and their empirical potential highlighted via two applications. The out-of-sample forecast comparisons are made using the *Superior Predictive Ability* test, which specifically guards against the perils of *data snooping*. Certain tentative conclusions are drawn regarding the relative forecasting ability of the different methods.

* *Correspondence Address*: Jose G. Clavel, Dpto. Metodos Cuantitativos, Facultad de Economia, Campus de Espinardo, 30100 Murcia, Spain. (email: jjgarvel@um.es). This work was partially supported by Fundación Séneca ref: 00924/EE3/04.

Keywords: mixed spectrum; autoregressive methods; eigenvalue methods; dynamic harmonic regression; data snooping; multiple forecast comparisons.

JEL classification: C22, C52, C53.

1 INTRODUCTION

Many series occurring in fields as diverse as seismology, astronomy, oceanography, acoustics, medicine and economics, exhibit strongly periodic behaviour. The early attempts to model such behaviour were undertaken by statisticians such as Schuster (1898), Slutsky (1937) and Yule (1927), who termed the problem as that of “detection of hidden periodicities”. The formal analytical treatment of this problem however, may be said to begin only in the 1950s, with the seminal contributions of Whittle (1952), Bartlett (1954), Moran (1953) and Grenader and Rosenblatt (1957). Several practical solution methods evolved in the wake of these contributions. In a broad taxonomy, approaches to studying periodic behaviour may be divided into two distinct groups.

i. Time Domain Methods such as

- (a) the well-known Census X-11 method and its several extensions (Findley, Monsell, Bell, Otto and Chen (1996))
- (b) ARIMA Decomposition methods (Hillmer, Bell and Tiao (1983), Gomez and Maravall (1996) etc.)
- (c) Optimal Regularization methods (Akaike (1980), Young (1991) etc.)
- (d) Stochastic State Space Methods (Harvey (1989), Koopmans (1993) etc.)

ii. Frequency Domain Methods comprising

- (a) Spectral Analysis (Priestley (1981), Marple (1987) etc.)
- (b) Wavelet Analysis (Percival and Walden (2000), Gençay, Selçuz and Whitcher (2002) etc.)

- (c) Harmonic Regression or Fixed Frequency Effects Models (Kay and Marple (1981), Hannan (1973), Chiu (1989) etc.)

In this paper, our focus is on the last category of the frequency domain methods viz. harmonic regression. In such models, an observed series $X(t)$, $t = 1, \dots, N$ is viewed as being composed of a *signal* (consisting of a sum of periodic terms) with a superimposed *coloured noise*.

$$X(t) = \sum_{p=1}^k (A_p \cos \omega_p t + B_p \sin \omega_p t) + u(t); \quad t = 1, 2, \dots, N \quad (1)$$

where $u(t)$, the *coloured noise* is a stationary process (i.e. capable of ARMA representation) and the A_i , B_i and ω_i are constant (but unknown) parameters to be estimated. The number of “harmonics” k may either be assumed known or unknown (though most of the methods that we discuss allow for it to be unknown). Additionally, we need to estimate the underlying parameters of the *coloured noise* process $u(t)$. The problem thus is fairly complicated. It needs to be noted that once k and the frequencies ω_i are determined, 1 becomes a regression and the remaining parameters are easily obtainable via linear regression methods. Estimating the parameters of $u(t)$ is usually undertaken after the sinusoidal signal is estimated. Thus, our discussion is focused on the estimation of k and the frequencies ω_i .

Our motivation towards concentrating on this group of models is three-fold. Firstly, many real world phenomena are dominated by a few strong periodic tendencies (the signal term in 1) contaminated by *coloured noise*. This happens for example, in oceanography (tidal phenomena), climatology (seasonal weather variations, rainfall cycles etc.), seismology (earthquakes and volcanic activity), and economics (seasonal and business cycle variations in series like industrial production, wholesale prices, money supply etc.). Such phenomena may thus be appropriately described by models of type (1). Secondly, the full potential of these models has yet to be realized in many disciplines. For example in economics, such models have seen little use (of the few exceptional applications, we may mention Hannan, Terrell and Tuckell (1970) and more recently Ghysels and Osborn (2001), and

Young and Pedregal (1999)). Finally, there has been virtually no evidence on the comparative forecasting ability of the several variants of the harmonic regression methodology which have been suggested in the literature.

The plan of our paper is as follows. Section 2 to 5 describes the salient features of four important categories of harmonic regression models. The out of sample forecast comparisons of the various methods are described in Section 6, with special attention to the recent emerging issue of *data snooping*. Finally, conclusions are gathered in Section 8.

The major methods suggested in the literature to deal with the problem (1), may be grouped under five headings:

1. Methods based on the periodogram or the DFT (discrete Fourier transform) (Whittle (1952), Walker (1973), Hannan (1973), Campbell and Walker (1977) etc.).
2. Mixed Spectrum methods (Priestley (1964, 1981), Bhansali(1979)).
3. Autoregressive methods including (i) the Prony method (Marple (1987), Candy (1988)) and (ii) Truong Van's (1990) Amplified Harmonics method.
4. Eigenvalue methods of which Pisarenko's Harmonic Decomposition method (Pisarenko (1973), Kay and Marple(1981)) may be taken as representative.
5. Dynamic Harmonic Regression method of Young, Pedregal and Tych (1999).

Of these, the first two are essentially frequency domain methods, the next two may be regarded as largely in the time domain, while the last uses results from both the time and frequency domains.

Thanks to the works of Hannan (1971,1973) and Walker (1973), the asymptotic properties of periodogram (and DFT) methods are well understood. Briefly, these methods yield fairly accurate estimates of the frequencies ω_i in (1) (with standard errors of order $N^{-3/2}$ for the periodogram method and N^{-1} for the DFT respectively), if the $u(t)$ are Gaussian. Their main

drawbacks (as noted in Bloomfield (1976), Rice and Rosenblatt (1988)) are two viz. their inability to distinguish frequencies separated by less than the reciprocal of the data length and the fact that the criterion function is likely to have several local minima, so that the final estimates are likely to have large biases, unless the iterations happen to commence at values close to the true frequencies.

Their main limitation is however theoretical in the sense that they fail to take cognizance of the distinction between the *discrete* and *continuous* spectrum for periodic processes such as those described by (1) —a distinction introduced into the literature by Priestley (1964). We therefore do not consider these methods here and pass on directly to the mixed spectrum methods.

2 MIXED SPECTRUM METHODS

2.1 Priestley's $P(\lambda)$ Test

Priestley (1964, 1981) showed that the spectrum $F(\omega)$ of processes such as (1) could be decomposed as

$$F(\omega) = F_1(\omega) + F_2(\omega)$$

where $F_1(\omega)$ is a *discrete* spectrum (corresponding to the trigonometric sum) and $F_2(\omega)$ is the *continuous* spectrum corresponding to the ARMA process $u(t)$. $F(\omega)$ is then called a *mixed spectrum*.

Reformulating (1) as

$$X(t) = \sum_{p=1}^k D_p \sin(\lambda_p t + \phi_p) + u(t) \quad (2)$$

(where D_p , λ_p , and k are unknown parameters, the ϕ_p , are independent and rectangularly distributed on $(-\pi, \pi)$ and $u(t)$ is a stationary linear process with a continuous spectrum). The following null hypothesis is tested

$$H_0 : D_p = 0, \quad p = 1, 2, \dots, k$$

against the alternative

$$H_1 : D_p \neq 0, \text{ for some } p, 1 \leq p \leq k,$$

Non-rejection of the null implies that $X(t)$ is a stationary ARMA process with a purely continuous spectrum. To test H_0 , Priestley (1981) proposes the $P(\lambda)$ test.

Let \hat{f}_m and \hat{f}_n denote two “window” estimates of the spectrum of $X(t)$, obtained using truncation points m and n respectively where $n > 2m$ (the window used for forming these estimates may be the same or different). We next put

$$P(\lambda) = \hat{f}_n(\lambda) - \hat{f}_m(\lambda) \quad (3)$$

at the Fourier frequencies $\lambda_j = \left(\frac{2\pi j}{N}\right)$; $j = 0, 1, \dots, [N/2]$.

If the D_p 's are not all zero, $P(\lambda)$ will have several well-defined peaks say $\omega_1 < \omega_2 < \dots < \omega_k$. These peaks are tested for significance (successively in order of occurrence) using the so-called \tilde{J}_q statistic (see Priestley (1981), p. 640). A notable feature of the test is that in the event of rejection of H_0 , it also suggests estimates of D_p , λ_p and k . Bhansali (1979) has suggested an important amendment to the $P(\lambda)$ test, which improves its performance significantly. We therefore proceed by using the test, with the Bhansali “correction” incorporated.

2.2 Model Estimation

Suppose by following the above procedure, we identify k harmonics at the frequencies ω_j ($j = 1, 2, \dots, k$). We then estimate the following model by OLS

$$X(t) = \sum_{p=1}^k \{A_p \cos(\hat{\omega}_p t) + B_p \sin(\hat{\omega}_p t)\} + u(t); \quad t = 1, 2, \dots, N \quad (4)$$

The fact that $u(t)$ may be correlated is not much of a cause for concern, since Durbin (1960) has shown that for harmonic regressions of the type (4), OLS estimates of A_p and B_p are asymptotically efficient.

We have already seen that the residual term $u(t)$ will have a *continuous spectrum*, with an ARMA representation. However, for forecasting purposes, an AR model is more convenient, and Bhansali (1979) shows how an autoregressive model may be fitted to $u(t)$ by a suitable lag selection criterion. His preference is for the FPE_α criterion developed in Bhansali and Downham (1977), according to which, the AR order of $u(t)$ is taken as the value of p (say p^*) which minimizes $FPE_\alpha(m)$, where

$$FPE_\alpha(p) = \hat{\sigma}_p^2 [1 + (\frac{\alpha p}{N})] \quad p = 0, 1, 2, \dots, L \quad (5)$$

where α is a positive constant, with suggested values (see Bhansali and Downham (1977)) in the range $(2, 4]$, and L is a suitably large positive integer.

Thus the Priestley-Bhansali method may be viewed as comprising the following three stages in fitting a *harmonic regression* or *fixed frequency effects* model to the data.

1. We first estimate the number of harmonics k using Priestley's $P(\lambda)$ test.
2. Next, we estimate the coefficients in the model (3) (using the value of r obtained in Stage 1) by the method suggested originally by Priestley (1981), with the Bhansali (1979) *correction* incorporated.
3. Finally, we fit an AR model to the residuals from the model estimated in Stage 2.

3 AUTOREGRESSIVE METHODS

The autoregressive methods derive their name from the fact that the sinusoidal part of (1) may be studied in terms of a homogenous AR equation, with coefficients uniquely related to the sinusoidal frequencies. The frequency estimation problem is thus transformed into the simpler AR estimation problem, which is amenable to solution by linear methods.

3.1 Prony Method

The Prony method was proposed in a now classic article by Gaspard Riche, Baron de Prony in 1795 (Prony, 1795). It is based on the observation that the deterministic trigonometric component of (1) satisfies a homogeneous autoregressive equation of order $2k$. Let $Y(t)$ denote the sinusoidal component of (1), so that (1) or (2) can be written as

$$X(t) = Y(t) + u(t) \quad (6)$$

with

$$Y(t) = \sum_{p=1}^k D_p \cos(\lambda_p t + \phi_p) \quad (7)$$

Let B denote the backward shift operator defined by $BY(t) = Y(t-1)$, and define the polynomial

$$A(B) = \sum_{p=1}^k (1 - z_p B)(1 - \bar{z}_p B) = \sum_{p=0}^{2k} a_p B^p \quad (8)$$

where

$$z_p = \exp(i\omega_p) \text{ and } \bar{z}_p = \exp(-i\omega_p) \quad (9)$$

It is easy to see (e.g. Candy (1988), p. 210–211) that (i) the a_p are uniquely determined by ω_p , and vice versa and (ii) the $(2k+1)$ coefficients a_p of the polynomial $A(B)$ are symmetric in the sense

$$a_0 = 1 \text{ and } a_{2k-p} = a_p; \quad p = 1, 2, \dots, k-1 \quad (10)$$

It can be shown further that the sinusoidal signal $Y(t)$ satisfies the following homogeneous AR equation of order $2k$

$$\sum_{p=0}^{2k} a_p Y(t-p) = 0 \quad (11)$$

The Prony method assumes that the number of harmonics k is known in advance and concentrates on estimation of the frequencies. By the above discussion, the frequencies can be determined from the zeros of the polynomial $A(B)$ once the coefficients a_p in (11) are known. The latter can be

obtained as the ordinary least squares estimate of the following regression which makes use of the fact of the symmetry of these coefficients noted in (10), (see Kay and Marple (1981), Li and Kedem (1992))

$$x = -\Gamma Q a + e \quad (12)$$

where $e(t) = A(B)u(t)$ and

$$x = \begin{bmatrix} X(2k+1) + X(1) \\ \dots \\ X(N) + X(N-2k) \end{bmatrix}; \quad e = \begin{bmatrix} e(2k+1) \\ \dots \\ e(N) \end{bmatrix};$$

$$\Gamma = \begin{bmatrix} X(2k) & \dots & X(2) \\ \dots & \dots & \dots \\ X(N-1) & \dots & X(N-2k+1) \end{bmatrix}; \quad Q = \begin{bmatrix} I & O \\ O^T & I \\ \bar{I} & O \end{bmatrix}$$

where I is $(k-1) \times (k-1)$ identity matrix, \bar{I} is $(k-1) \times (k-1)$ reverse permutation matrix with 1's on the anti-diagonal and 0's elsewhere, and O is the zero vector of dimension $(k-1)$.

From (12), the Prony estimator may be written as

$$\hat{a} = -(Q^T \Gamma^T \Gamma Q)^{-1} Q^T \Gamma^T x \quad (13)$$

While the Prony method is computationally convenient, the resultant estimates can be biased. Several methods have been suggested in the literature to overcome this bias including using higher order AR models in (11) (see Lang and McClellan (1980)), contraction mapping methods (He and Kedem (1989), Yakowitz (1991)) and parametric filtering (Li and Kedem(1992)). However, most of these methods are computationally intensive, and hence the scope of their applications is, at the moment, somewhat restrictive. Hence we do not consider these extensions here. The other limitation of the Prony method refers to the fact that it presumes knowledge of k (the number of harmonics in the data). This limitation is usually overcome via resort to a two-stage procedure-estimating k in the first stage by using methods such as those suggested by Chiu (1989), Quinn and Thomson (1991) etc., and then applying the Prony method for this fixed k . The method suggested by Chiu (1989) is particularly simple to apply, and we consider this below.

3.2 Estimating the Number of Harmonics (Chiu's Method)

Chiu (1989) suggests two alternate statistics to test the null of

H_0 : zero harmonics in $X(t)$

against the alternative

H_1 : r harmonics in $X(t)$

The statistics are defined as follows

$$U(r) = \frac{I_{n-r+1}}{\sum_{i=1}^n I_i} \quad (14)$$

$$V(r) = \frac{I_{n-r+1}}{\sum_{i=1}^{n-r} I_i} \quad (15)$$

Where I_m denotes the periodogram of $X(t)$, ($t = 1, 2, \dots, N$) ordinate at the m -th Fourier frequency $\omega_m = (\frac{2\pi m}{N})$.

Chiu (1989) derives the asymptotic distribution of $U(\cdot)$ and $V(\cdot)$ for testing H_0 as follows:

Define

$$Z_1(r) = nU(r) - \ln(n - r + 1) \quad (16)$$

$$Z_2(r) = c(n - r)V(r) - \ln(n - r + 1) \quad (17)$$

where

$$c = 1 + \frac{r[\ln(r/n)]}{(n - r)} \quad (18)$$

Also let

$$P_i(r) = \exp\{-\exp[-Z_i(r)]\} \sum_{j=0}^{r-1} \exp\{-jZ_i(r)/j!\}; \quad i = 1, 2 \quad (19)$$

If the selected level of significance is α , we reject H_0 in favour of H_1 whenever

$$P_i(r) > (1 - \alpha); \quad i = 1, 2 \quad (20)$$

3.3 Method of Amplified Harmonics

This method was introduced into the literature by Truong-Van (1990). The underlying model is taken to be (2) with $u(t)$ having a stationary ARMA representation. The number of data points is taken to be N as before. The logic of the method relies on the notion of amplification of harmonics at various frequencies. Truong-Van (op. cit.) defines for each frequency ω_j , a process $\xi_t(\omega_j^*)$ by the the following recursion where ω_j^* is near to ω_j

$$\xi_t(\omega_j^*) = 2 \cos(\omega_j^*) \xi_{t-1}(\omega_j^*) + \xi_{t-2}(\omega_j^*) + X(t); \quad t = 1, 2, \dots, N \quad (21)$$

with $\xi_0 = \xi_{-1} = 0$

It is then shown that the process $\xi_t(\omega_j^*)$ amplifies the harmonic of frequency ω_j , selectively relative to the others.

Truong-Van (1990, Theorem 3) is then led to demonstrate that among the harmonic amplifiers $\xi_t(\omega_j^*)$ of ω_j , there exists an amplifier $\xi_t(\hat{\omega}_j)$ s.t.

$$\sum_{t=2}^N \xi_{t-1}(\hat{\omega}_j) X(t) = 0 \quad (22)$$

i.e. $\xi_{t-1}(\hat{\omega}_j)$ is orthogonal to $X(t)$.

A suggested estimate of ω_j is then $\hat{\omega}_j$. These estimates are strongly consistent and asymptotically normal (see Hannan(1973) and Truong-Van (1990, Theorem 4)). From the computational point of view, the following result is important (Truong-Van, op. cit. Equation (9)).

Proposition

Consider the problem of estimating α in the following regression by OLS,

$$\xi_t(\omega_j^*) + \xi_{t-2}(\omega_j^*) = 2\alpha \xi_{t-1}(\omega_j^*) + \epsilon_t \quad (23)$$

where ω_j^* is near to ω_j .

Let $e_t(\omega_j^*)$ denote the OLS residuals of (23). Then $\hat{\omega}_j$ is the solution to the following minimisation problem:

$$\min_{\omega_j^* \in V(\omega_j)} \sum_{t=1}^N \{e_t(\omega_j^*) - X(t)\}^2 \quad (24)$$

Note: The neighbourhood $V(\omega_j)$ of ω_j is defined as follows. Let $u(t)$ in (2) have the ARMA representation

$$\Phi(B)u(t) = \Theta(B)a(t) \quad (25)$$

where $a(t)$ are i.i.d. with mean 0 and variance σ^2 . Let $(2\pi/\sigma^2)f(\omega)$ be the spectrum of $u(t)$, then $V(\omega_j)$ is defined as

$$V(\omega_j) = \frac{24(\omega_j)\sigma^2}{D_j^2} \quad (26)$$

where D_j^2 are as defined in (2).

The solution of this somewhat intricate problem can proceed along either of the two lines suggested by Truong-Van (1990)

1. signal orthogonal amplifiers by double amplification of harmonics (SODAH)
2. recursive least squares on amplified harmonics (RLSOAH)

The SODAH algorithm requires starting values of ω_j^* fairly close to the true values ω_j , while RLISOAH is more robust to the choice of initial values. We therefore resort to the RLISOAH algorithm. This algorithm will lead to estimates $\omega_1, \omega_2, \dots, \omega_k$, of the harmonics— the algorithm also identifying k , the number of harmonics. This latter fact constitutes a major advantage over the Prony method discussed above. The estimation of model (2) is now straightforward. We estimate the following equation by OLS

$$X(t) = \sum_{p=1}^k \{A_p \cos(t\hat{\omega}_p) + B_p \sin(t\hat{\omega}_p)\} + u(t) \quad (27)$$

The residuals $u(t)$ from this model can then be used to identify an ARMA(p,q) model for $u(t)$ in the standard fashion (one should possibly test $u(t)$ for stationarity to establish the reliability of the estimates $\hat{\omega}_1, \hat{\omega}_2, \dots, \hat{\omega}_k$).

4 EIGENVALUE METHODS

Eigenvalue methods are so called because they obtain the frequencies ω_p in (1) or (2) as the roots of a polynomial of degree k , whose coefficients are

derived from the eigenvector of a particular matrix B . The Prony method discussed above may also be regarded as an eigenvalue method (see e.g. Smyth (2000)), though it is now conventionally viewed as an autoregressive method. Among the eigenvalue methods, possibly the best known is Pisarenko's *harmonic decomposition method*, the essential features of which we sketch below.

4.1 Pisarenko's Harmonic Decomposition Method

Pisarenko's (1973) method proceeds by considering the sinusoidal model (2). As originally suggested, it suffers from two major limitations viz. the assumption that $u(t)$ is a white noise (rather than a general covariance stationary) process and the assumption that k , the number of sinusoids is known. As we shall see below, both these restrictions have been removed in later refinements of the method. But for illustrative purposes, and to keep the analytics simple, we will proceed within the original Pisarenko formulation. Consider then the model (2), with k known and $u(t)$ a white noise. Let $R_{XX}(p)$ denote the autocovariance of the process $X(t)$ at lag k . Then a straightforward application of the Wiener-Khintchine theorem (see Cadzow (1982)) shows that

$$R_{XX}(m) = \sum_{p=1}^k \left(\frac{D_p^2}{2} \right) \cos(k\omega_p) + \sigma_u^2 \delta(m); \quad m = 0, 1, \dots, k-1 \quad (28)$$

where $\sigma_u^2 = \text{var}(u(t))$ and $\delta(m)$ is the Kronecker δ .

Using a well known result from linear systems theory (see Kailath (1980)), the sinusoidal term in (2), which we have earlier denoted as $Y(t)$, satisfied the following difference equation of order $2k$

$$Y(t) = - \sum_{p=1}^{2k} \alpha_p Y(t-p) \quad (29)$$

where the constants α_p are uniquely related to the frequencies ω_p in (2). Using (29) in (2) (or (6)) we get

$$X(t) = - \sum_{p=1}^{2k} \alpha_p X(t-p) + \sum_{p=0}^{2k} \alpha_p u(t-p) \quad (30)$$

with $\alpha_0 = 1$.

Define

$$\underline{X}(t) = \begin{bmatrix} X(t) \\ \dots \\ X(t-2K) \end{bmatrix}; \quad \underline{Y}(t) = \begin{bmatrix} Y(t) \\ \dots \\ Y(t-2k) \end{bmatrix};$$

$$\underline{U}(t) = \begin{bmatrix} u(t) \\ \dots \\ u(t-2k) \end{bmatrix}; \quad \underline{\alpha}(t) = \begin{bmatrix} \alpha_0 \\ \dots \\ \alpha_{2k} \end{bmatrix}$$

Then (6) may be written as

$$\underline{X}(t) = \underline{Y}(t) + \underline{U}(t) \quad (31)$$

and (30) assumes the form

$$\underline{X}^T \underline{\alpha} = \underline{U}^T \underline{\alpha} \quad (32)$$

From the above two equations, the following relation can be easily derived

$$\left[R_{XX} - \sigma_u^2 I \right] \underline{\alpha} = 0 \quad (33)$$

with R_{XX} a Toeplitz matrix (see Brillinger (1975), p.72 for a definition). Thus, in view of (33), the parameter vector (of the autoregressive coefficients in (30)) is the eigenvector associated with the eigenvalue σ_u^2 normalized on $\alpha_0 = 1$. This result of itself is of little use, since the value σ_u^2 is not known. However, Pisarenko (1973) shows that if $u(t)$ is white noise, then σ_u^2 corresponds to the *minimum eigenvalue* of R_{XX} . The algorithm for calculation of the minimum eigenvalue of R_{XX} is based on the classical *power method* of matrix algebra (see Stewart (1973)), based on the iteration

$$R_{XX} \underline{\alpha}^{(k)} = \underline{\alpha}^{(k-1)} \quad (34)$$

where $\underline{\alpha}^{(k)}$ is the estimate of α at the k -th iteration, starting with an initial guess $\underline{\alpha}^{(0)}$. Once the iterations have converged and an estimate $\underline{\alpha}^{(*)}$ of $\underline{\alpha}$ obtained, the minimum eigenvalue λ_{\min} of R_{XX} is given by the Rayleigh quotient (see Magnus and Neudecker (1991), p. 203)

$$\lambda_{\min} = \frac{\underline{\alpha}^T R_{XX} \underline{\alpha}}{\underline{\alpha}^T \underline{\alpha}} \quad (35)$$

Define the same matrix polynomial $A(B)$ as in (8) with the autoregressive coefficients corresponding to the elements of the vector $\underline{\alpha}^*$ (and with $\alpha_0 = 1$). The symmetry condition (10) also applies, since the matrix R_{XX} is Toeplitz. The frequencies ω_p can now be obtained from the zeros of the polynomial $A(B)$ exactly as in the Prony case. Thus the essential difference between the Prony and Pisarenko estimates lies in the way the coefficients of $A(B)$ are estimated (by Least Squares in the Prony case and by eigenanalysis in the Pisarenko case). However as shown by Smyth (2000) the Pisarenko estimates are also biased.

As mentioned earlier, the Pisarenko method suffers from two major restrictions, viz. that the number of harmonics k is assumed known, and the noise $u(t)$ is assumed to be white. The first of these restrictions can be removed by adopting a two-stage procedure as suggested in Subsection 3.2. The second restriction is more difficult to deal with, and several methods have been suggested to generalize the method to situations where the variance covariance matrix of $u(t)$ need not be an identity matrix (see in particular Sakai (1984), Stoica and Nehorai (1988), Kundu and Kannan (1997) etc.).

5 DYNAMIC HARMONIC REGRESSION (DHR) METHOD

The dynamic harmonic regression method (DHR) attempts to take the model (1) (or its equivalent formulation (2)), further in the direction of reality by allowing for the possibility that the amplitudes and the phases of the underlying sinusoids (i.e. D_p and ϕ_p in (2)) may vary over time. The method as developed by Young and Pedregal (1999), Young et al (1999) etc. proceeds within the framework of the Unobserved Components (UC) model, frequently used in the analysis of economic time series, viz.

$$X(t) = T(t) + C(t) + S(t) + e(t) \quad (36)$$

Where $T(t)$, $C(t)$ and $S(t)$ are the trend, cyclical and seasonal components respectively, and $e(t)$ the irregular component is assumed to be a Gaussian white noise with constant variance σ_e^2 .

Suppose that the number of cyclical sinusoids is k_1 , and the number of seasonal sinusoids is k_2 , then (36) can be written analogously to (1) as

$$X(t) = \sum_{p=0}^k (A_p(t) \cos \omega_p t + B_p(t) \sin \omega_p t) + e(t); \quad (37)$$

$$(k = 1 + k_1 + k_2); \quad (t = 1, 2, \dots, N)$$

There are three major differences between (1) and (37):

1. (37) includes a zero frequency term $\omega_0 = 0$ corresponding to the trend $T(t)$
2. The disturbance term $e(t)$ in (37) is restricted to be white noise.
3. The parameters are allowed to vary with time in (37) whereas they are constant in (1). This of course, means that both the amplitude and phase of the cyclical and seasonal components (as well as the level and slope of the trend) are allowed to change over time.

In view of item n. 3, the model (37) may be viewed as a generalization of (1). The evolution of each of the $2k$ parameters in (37) is characterized by two stochastic variables -the level $l(t)$ and the slope $d(t)$. For the i -th parameter we denote these by a two-dimensional stochastic vector

$$\theta_i(t) = \begin{bmatrix} l_i(t) \\ d_i(t) \end{bmatrix}; i = 1, \dots, 2k \quad (38)$$

which is assumed to evolve as a Generalized Random Walk (GRW) process (see Jakeman and Young (1984) for a definition and discussion of the properties of a GRW process) evolving as follows:

$$\theta_i(t) = F_i \theta_i(t-1) + G_i \eta_i(t); \quad i = 1, 2, \dots, 2k \quad (39)$$

with

$$F_i = \begin{bmatrix} \alpha & \beta \\ 0 & \gamma \end{bmatrix}; \quad G_i = \begin{bmatrix} \gamma & 0 \\ 0 & 1 \end{bmatrix}$$

We now aggregate all the subsystem matrices described by (39) into a standard state space (SS) format (see Harvey (1989), or Koopmans, Harvey, Doornik and Shephard (1995))

Observation Equation:

$$X(t) = H(t)\Theta(t) + e(t); \quad (40)$$

State Equation:

$$\theta(t) = F\Theta(t-1) + G\eta(t) \quad (41)$$

Where F is a $(4k \times 4k)$ block diagonal matrix with each block defined by a matrix F_i from the subsystem (39); G is also $(4k \times 4k)$ and obtained similarly from the matrices G_i in (39); $H(t)$ is an appropriately defined $(1 \times 4k)$ vector which relates the scalar observation $X(t)$ to the state variables defined by (39), so that it represents the UC model (37); $\Theta(t), \eta(t)$ are $(4k \times 1)$ vectors obtained by stacking up the θ_i, η_i vectors from (39). Finally the $\eta_i(t)$ are assumed to be individually white noise and uncorrelated with the observation noise $e(t)$ in (40). We denote the covariance matrix of $\eta(t)$ by Θ .

As is well known, the system (40)–(41) admits of a recursive solution, either using the forward pass filtering algorithm as originally suggested by Kalman (1960), or the backward pass smoothing algorithm of Bryson and Ho (1960). However, the solution requires knowledge of the matrix Θ (the covariance matrix of the system disturbances) as well as of the variance of the observation noise σ_e^2 . Obtaining estimates of these parameters constitutes the major innovation of the DHR method.

The DHR method is applicable for a variety of GRW processes such as

1. the integrated random walk (IRW) with $\alpha = \beta = \gamma = 1; \delta = 0$.
2. the scalar random walk (RW) with $\alpha = \beta = \delta = 0; \gamma = 1$
3. the smoothed random walk (SRW) with $0 \prec \alpha \prec 1; \beta = \gamma = 1$ and $\delta = 0$
4. local linear trend (LLT) with $\alpha = \beta = \gamma = 1; \delta = 1$ and
5. damped trend (DT) with $\alpha = \beta = \delta = 1; 0 \prec \gamma \prec 1$.

For the sake of specificity, we illustrate the DHR method when the underlying GRW process in (39) is an integrated random walk (IRW) process. The pseudo-spectrum $f_{\omega_j}(\omega)$ of a sinusoid

$$\xi(t) = a(t) \sin(\omega_j t) + b(t) \cos(\omega_j t) \quad (42)$$

in which both $a(t)$ and $b(t)$ follow IRW processes with a common variance $\sigma_{\omega_j}^2$ is derivable as (see Young et al (1999))

$$f_{\omega_j}(\omega) = \sigma_{\omega_j}^2 S(\omega, \omega_j) \quad (43)$$

where

$$S(\omega, \omega_j) = \frac{1}{2\pi} \left[\frac{1}{4\{1 - \cos(\omega - \omega_j)\}^2} + \frac{1}{4\{1 - \cos(\omega + \omega_j)\}^2} \right] \quad (44)$$

With this notation established, the pseudo-spectrum of the full DHR model (37) can be written as

$$f_x(\omega, \underline{\sigma}^2) = \sum_{p=0}^k \sigma_{\omega_p}^2 S(\omega, \omega_p) + \frac{\sigma_e^2}{2\pi} \quad (45)$$

where

$$\underline{\sigma}^2 = [\sigma_e^2, \sigma_{\omega_0}^2, \sigma_{\omega_1}^2, \dots, \sigma_{\omega_k}^2]$$

The DHR estimation now proceeds as follows. Given the observed process $X(t)$, $t = 1, 2, \dots, N$, we estimate an autoregressive spectrum AR(m) for $X(t)$ where the order m is identified by a lag selection procedure such as the AIC (the AR spectrum is explained, for example, in Gardner (1988), p. 266). *The significant peaks in this spectrum gives us an estimate of k (the number of sinusoids) as well as the corresponding frequencies ω_p , $p = 1, 2, \dots, k$.* In the next stage, we evaluate the quantity $f_x(\omega, \underline{\sigma}^2)$ at select frequencies $\omega = \omega_j$, $j = 1, 2, \dots, T - 1$ in the interval (0,0.5) as the empirical spectrum of $X(t)$, while the expression $\sum_{p=0}^k S(\omega, \omega_p)$ is also evaluated via (44) (with $\omega = \omega_j$, $j = 1, 2, \dots, T - 1$ and ω_p , $p = 1, 2, \dots, k$ from the first stage). The parameter vector $\underline{\sigma}^2$ can now be obtained as the coefficients of a linear

regression of $f_X(\omega, \underline{\sigma}^2)$ on the term $\sum_{p=0}^k S(\omega, \omega_p)$ and an intercept. This, of course, corresponds to minimizing

$$J(f_X, \hat{f}_X) = \sum_{j=0}^{T-1} \left[f_X(\omega_j, \underline{\sigma}^2) - \hat{f}_X(\omega_j, \hat{\underline{\sigma}}^2) \right]^2 \quad (46)$$

where

$$\hat{f}_X(\omega, \hat{\underline{\sigma}}^2) = \sum_{p=0}^k \hat{\sigma}_{\omega_p}^2 S(\omega, \omega_p) + \frac{\hat{\sigma}_e^2}{2\pi} \quad (47)$$

In practical applications, better results can be obtained by minimizing a slightly different criterion

$$(f_X, \hat{f}_X) = \sum_{j=0}^{T-1} \left[\log\{f_X(\omega_j, \underline{\sigma}^2)\} - \log\{\hat{f}_X(\omega_j, \hat{\underline{\sigma}}^2)\} \right]^2 \quad (48)$$

Once the parameter vector $\hat{\underline{\sigma}}^2$ is obtained, the system (40)–(41) is completely solved via standard recursive state space methods.

The above derivation applies for the specific case where the parameters of the various sinusoids follow IRW processes. Appropriate modifications are in order when other patterns of behaviour (such as SRW, LLT or DT) are assumed.

6 MULTIPLE FORECAST COMPARISONS AND DATA SNOOPING

We have in effect discussed five different methods for estimating models with strong periodic components viz.

1. Mixed spectrum method of Priestley (Method I)
2. Prony's autoregressive method (Method II)
3. Truong-Van's method of amplified harmonics (Method III)
4. Pisarenko's harmonic decomposition method (Method IV)
5. Dynamic harmonic regression method (Method V)

It is our intention to attempt an assessment of the comparative forecasting abilities of the various models discussed above via standard empirical illustrations. The issue of devising formal procedures for comparing and evaluating forecasts made by competing methods has, of late, been receiving considerable attention. Following the pioneering work of Diebold and Mariano (1995) and West (1996), multiple forecast comparisons are now recognized as constituting a non-standard inference problem. White (2000) in particular, cautions against *data snooping*, which refers to the possibility of any given forecasting method performing well by mere chance, than through any intrinsic merit, in a situation where the same data set is used more than once for purposes of inference or prediction. He also proposes a so-called *reality check* test T^{RC} to test the hypothesis that the best model encountered in a specification search has no predictive superiority over any given benchmark model. In a later contribution, however, Hansen (2001) puts forth an alternative test statistic T^{SPA} (superior predictive ability statistic), which while not uniformly more powerful than T^{RC} , performs better in most empirically relevant situations, and is less sensitive to the inclusion of irrelevant or poor alternatives in the specification set. We therefore propose to use Hansen's statistic in the evaluation of the forecasting abilities of the seven alternative methods constituting our specification set.

6.1 Superior Predictive Ability Test (Hansen)

Let the specification set comprise m models indexed by $k = 1, 2, \dots, m$, with $k = 0$ denoting the benchmark model. We denote the 1-period ahead forecast for period t (made at period $(t - 1)$) using the $k - th$ method as $F_X^{(k)}(t)$ and let $L(X(t), F_X^{(k)}(t))$ denote a suitable loss function. The performance of $k - th$ method relative to the benchmark (over our forecast sample) may now be defined as

$$\delta^{(k)}(t) = L\left(X(t), F_X^{(0)}(t)\right) - L\left(X(t), F_X^{(k)}(t)\right); \quad t = n + 1, \dots, N \quad (49)$$

Let the performance vector of the m methods considered (relative to the benchmark) at time t be denoted as

$$\Psi(t) = \left(\delta^{(1)}(t), \delta^{(2)}(t), \dots, \delta^{(m)}(t)\right); \quad t = n + 1, \dots, N \quad (50)$$

If $\mu = E(\Psi(t))$ exists, then the null hypothesis may be set up as

$$H_0 : \mu \leq 0 \quad (51)$$

i.e. the benchmark model is not inferior to any of the models in the specification set.

Under certain fairly general assumptions, the expectation vector μ in (51) is well defined and $\sqrt{(N-n)}(\bar{\Psi} - \mu)$ converges in distribution to the multivariate normal density $N(0, \Omega)$, where

$$\bar{\Psi} = (\bar{\Psi}^{(1)}, \bar{\Psi}^{(2)}, \dots, \bar{\Psi}^{(m)}) = (N-n)^{-1} \sum_{t=n+1}^N \Psi(t) \quad (52)$$

and Ω is the covariance matrix defined by

$$\Omega = \lim_{n \rightarrow \infty} E \left[(N-n)(\bar{\Psi} - \mu)(\bar{\Psi} - \mu)^T \right] \quad (53)$$

Hansen (2001) defines the studentized statistic

$$T^{SPA} = \max \left[\max_{k=1,2,\dots,m} \frac{\sqrt{n}\bar{\Psi}^{(k)}}{\hat{\omega}^{(k)}}, 0 \right] \quad (54)$$

where $\hat{\omega}^{(k)}$ is a consistent estimator of

$$\hat{\omega}^{(k)} = \left[\text{var} \left(\sqrt{(N-n)}\bar{\Psi}^{(k)} \right) \right]^{0.5} \quad (55)$$

and $\hat{\mu}^{(k)}$ is a consistent estimate of

$$\mu^{(k)} = E \left[\bar{\Psi}^{(k)} \right] \quad (56)$$

Since a precise distribution of Ω is unavailable, operationalizing the test procedure has to proceed via bootstrapping methods. In our paper, we propose to follow earlier empirical studies (Sullivan, Timmerman and White (1999), Hansen and Lunde (2005)) in resorting to the stationary bootstrap of Politis and Romano (1994). In the present context, the application of this method requires bootstrap sampling from the vector $(\Psi(n+1), \dots, \Psi(N))$ where $\Psi(t)$, $(t = n+1, \dots, N)$ is as given by (50). Let the typical bootstrap

vector be denoted by $\Psi^{*b}(t)$ and denote by $\hat{\omega}^{*(k)}$ the bootstrap estimate of (55), with $\hat{\mu}^{*(k)}$ the corresponding bootstrap estimate of (56). The distribution of T^{SPA} in (54) may be approximated by the empirical distribution of the bootstrap estimate

$$T^{*b,SPA} = \max_{k=1,\dots,m} \sqrt{n} \left(\bar{\Psi}^{*b,(k)} - \hat{\mu}^{*(k)} \right); \quad b = 1, 2, \dots, B \quad (57)$$

where B is the total number of bootstrapping samples and $\bar{\Psi}^{*b,(k)}$ is the k -th component of the vector $\bar{\Psi}^{*b}$ defined as

$$\bar{\Psi}^{*b} = \left(\bar{\Psi}^{*b,(1)}, \dots, \bar{\Psi}^{*b,(m)} \right) = (N - n)^{-1} \sum_{t=n+1}^N \Psi^{*b}(t); \quad b = 1, 2, \dots, B \quad (58)$$

Let $T^{O,SPA}$ denote the observed value of T^{SPA} (computed from (54) using the bootstrap estimate $\hat{\omega}^{*(k)}$), then the null hypothesis (51) is rejected if the p -value of the statistic

$$\hat{p}^{SPA} = B^{-1} \sum_{b=1}^B I \left[T^{*b,SPA} \succ T^{O,SPA} \right] \quad (59)$$

is small (i.e. less than 0.01 or 0.05, depending on the chosen level of significance), and where $I[\cdot]$ is once again the indicator function.

6.2 Statistical Loss Functions

A number of loss functions for forecast evaluation have been suggested in the literature. As, *a priori*, there seems to be little ground for choosing any particular loss function, we select the following two which have seen extensive use by earlier researchers in the field.

1. RMSE (Root Mean Square Error) =

$$= \sqrt{\frac{1}{N - n} \sum_{t=n+1}^N e^2(t/t - 1)}$$

2. MAPE (Mean Absolute Percentage Error) =

$$= \left(\frac{1}{N - n} \right) \sum_{t=n+1}^N \left[\frac{1}{|X(t)|} \right] |e(t/t - 1)|$$

3. MAXAE (Maximum Absolute Error) =

$$= \max\{e(t/t - 1)\}_{t=n+1}^N$$

In the above $e(t/t - 1)$ is the one-period ahead forecast error associated with the model being tested, $X(t)$ is the actual observation at time t , where n is the number of observations in the training sample, and $(N - n)$ the number of observations in the forecasting sample.

Additionally, very often, especially for economic time series, forecasting the direction of change becomes an important consideration. We therefore introduce an additional accuracy measure viz. D (the *proportion* of times the model correctly forecasts the sign of the series). The loss function in this case may be defined (following Pesaran and Timmerman (1992)) as

4. LS (Loss Function for Sign Checking) =

$$= 1 - \left(\frac{1}{N - n} \right) \sum_{t=n+1}^N Z^{(k)}(t); \quad k = 1, 2, \dots, m$$

$$Z^{(k)}(t) = 1, \text{ if } X(t)F_X^{(k)}(t) \succ 0; \quad Z^{(k)}(t) = 0, \text{ otherwise;} \\ t = n + 1, \dots, N$$

with, as before, $X(t)$ denoting the actual value of the series at time t and $F_X^{(k)}(t)$ the one-period ahead forecast made by the $k - th$ method. Our loss function LS thus measures the proportion of times a given method makes an incorrect sign prediction.

7 EMPIRICAL ILLUSTRATION

7.1 Sunspot Data

The first pertains to the standard sunspot data set, Wolfer’s Annual Sunspot Series (1701–1955) (see Waldheimer (1961)), often used in time series research. Since the series indicates presence of a unit root (on the basis of the Dickey-Fuller and Phillips-Perron tests), we apply our five methods discussed above, to the first differences of the series. Of the 254 observations at our disposal, we use 230 as the *training sample* (i.e. to obtain model parameter estimates), while the 24 observations at the end are used as the *forecasting sample* (i.e. to test the out-of-sample forecasting performance of each method). For the latter, only one period ahead forecasts are considered, which are made on the basis of all the information available at the time of making the forecast, except that the model parameters are fixed at the values obtained over the training sample. In terms of the terminology introduced above, $N = 254$ while $n = 230$.

In this case, the number of harmonics identified for the different methods were as follows:

Method I (Priestley-Bhansali Method):	5
Method II (based on Chiu’s Method):	6
Method III (Truong-Van Method):	4
Method IV (based on Chiu’s Method):	6
Method V (DHR Method):	5

The relevant results are summarised in Table A.1.

Methods I and V both suggest 5 cycles in the data, with periods ranging between 10.05 years to 91.39 years (for Method I) and 9.21 years to 72.77 years (for Method V); Chiu’s method identifies 6 cycles and using this estimate for Method II we located cycles ranging from 2.69 years to 85.37 years, (with the corresponding range for Method IV being 3.12 to about 62.8 years), while Method III yields cycles in the range 2.02 years to 52.58 years. Thus all five methods indicate that the sunspot data is characterized by about 4 to 6 cycles, ranging in frequency from 2 to 91 years. Using the

frequency estimates, obtained for each Method, we proceeded ahead to estimate the various other parameters (amplitudes and phases of the sinusoids as well as the variance of noise terms) as per the procedures outlined above for each of the five methods. After computing the out-of-sample forecasts, we are in a position to evaluate our five alternative forecasting. For this purpose, we resort to Hansen’s *Superior Predictive Ability test* discussed above. Following standard practice, we use the Random Walk model (denoted as Method VI) as an additional benchmark.

Application of Hansen’s *superior predictive ability test* necessitates much care in view of the specialized nature of the null hypothesis, in which a set of models is compared to a given benchmark. Since our interest centers around getting an idea about the relative forecasting ability of the models, we follow the strategy adopted by Hansen and Lunde (2005) of calculating the statistic \hat{p}^{SPA} successively using each of our five models as benchmarks and denote as \hat{p}^{SPA} ; $i = I, II, \dots, VI$, the value of Hansen’s statistic when the i – th model is being used as the benchmark. We take the size of the bootstrapping sample as 10,000 in each case (i.e. $B = 10,000$). The bootstrapping is done via the *stationary bootstrap* of Politis and Romano (1994) (the implementation procedure is described in detail in Nachane (2001)). This statistic is calculated for each of each of our four chosen loss functions (viz. RMSE, MAPE, MAXAE and LS) and for Methods I to VI. The relevant results for the Sunspots example are presented in Table A.2.

Examination of Table A.2. reveals that the null hypothesis (51) is rejected only for the random walk forecasts for three loss functions (viz. RMSE, MAPE and LS). Even for the remaining loss function MAXAE, there is only acceptance of the null at the margin. Thus we may confidently assert that all the harmonic regression models considered here comfortably outperform the naïve forecaster (the only exception is for the Prony method II, which works out to be the worst on one criterion viz. MAXAE). Only slightly better than the random walk is the Prony method (II). Of the other methods, Methods I and III are at the top followed by Methods V and IV respectively. That the performance of the DHR method (Method V), explicitly allowing for time variation in the coefficients, is somewhat inferior to the Priestley-

Bhansali and Truong-Van methods (Methods I and III) calls for comment. There could be two possible reasons for this. One relates to the fact that in the frequency estimation, the DHR method proceeds as if the spectrum were *continuous* rather than *mixed*, as is being done in Method I. This amounts to ignoring an important distinction feature of deterministic sinusoidal variation. Secondly, Method III only needs to assume the underlying series to be *locally stationary* and by making this simplifying assumption this method may be able to capture features of the data which are evolving slowly over time, as is likely to be the case with many series in natural sciences. The DHR method by assuming that the amplitudes and phases of the sinusoids are moving as per a GRW (generalized random walk) may be failing to account for some of the systematic but slowly evolving features of the sunspot series.

7.2 Index of Industrial Production (Spain)

Our next illustration pertains to an economic time series viz. the monthly IIP (Index of Industrial Production) in Spain from Jan 1985 to March 2006, a total of 255 observations. Once again the data were taken in first difference form, as both the Dickey-Fuller and Phillips-Perron tests indicated presence of unit roots.

For Spanish IIP , the number of harmonics identified by each method is presented below:

Method I (Priestley-Bhansali Method):	6
Method II (based on Chiu's Method):	6
Method III (Truong-Van Method):	3
Method IV (based on Chiu's Method):	6
Method V (DHR Method):	4

We summarise the results in Table A.3. The longest cycle in the data seems to be 10.5 years as identified by Method III, whereas the shortest cycle (detected by Method II) is about 2.4 months, very close to the Nyquist frequency. Table A.3. presents the sinusoidal frequencies estimated by each of our five models. Once again, we compute the Hansen \hat{p}^{SPA} statistic for

each of each of our four chosen loss functions (viz. RMSE, MAPE, MAXAE and LS) and for our five harmonic regression methods as well as for Method VI (random walk). The relevant results for the Spanish IIP data set are presented in Table A.4.

Here too, all the methods consistently out- perform the random walk (though Method II (Prony) is once again an exception). Method IV (Pisarenko) performs somewhat unremarkably, being only better than the Prony method and the random walk. Here the DHR method (Method V) shows its strength in forecasting economic time series, over the other methods. This, of course, is hardly surprising, since the DHR method is being suggested specifically in the context of the UC (Unobserved Components) model, which is usually postulated for economic series, its applicability in other fields being open to question.

8 CONCLUSIONS

The occurrence of periodicities is a characteristic feature of time series in both natural and social sciences. However, unraveling such periodic features from observed records is a difficult task, which has attracted a great deal of attention over the years. There have emerged in the statistical literature various approaches to deal with this problem, both in the time and frequency domains. One important group of methods viz. those based on harmonic regression, are the focus of our paper. We consider five specific methods in this class (see Section 1 for a detailed description). It should be noted that a common feature of four of the methods (Methods I to IV) considered in the paper, is an attempt to preserve as far as possible the stationarity assumption, and focus on the nonlinear aspects via a trigonometric approximating function in the standard spirit of Fourier analysis. Hence these approaches typically adopt a two-stage procedure in which firstly any non-stationarity, apparent in the original data, is sought to be removed, and the transformed series is modeled by alternative nonlinear trigonometric models. In practice, it would be extremely difficult to disentangle the separate effects of nonstationarity and nonlinearity, given that most of the tests for nonlinea-

rity are crucially dependent on the assumption of stationarity. The typical null being that the series under consideration is both linear and stationary, a rejection could be either in the direction of nonlinearity, nonstationarity or both. Recent attempts to separate nonlinearity and nonstationarity via so-called *phase scrambled bootstrap methods* (Aparicio (1998), Kugimutzis (2001) and Barnett and Wolff (2005)) have not been very successful. Method V attempts to meet this challenging task by attempting to model both the nonlinear and possibly nonstationary features of the data simultaneously, by making the coefficients time varying in a deterministic fashion. It is as well to remember that none of the five methods can claim to deal with all types of nonstationarity likely to be encountered in practice. In particular, the important type of nonstationarity arising from the presence of unit roots cannot be accommodated directly. A pragmatic approach would then be to remove any unit roots indicated by statistical tests, and then apply these methods to the residual component of the series. This approach may not be fully satisfying, but none other seems at the moment, available.

After a concise review of the alternative methods, we consider their application to two empirical illustrations, one from astronomy (sunspots data), the other from economics (Spanish IIP). These two empirical examples seem to indicate that the methods yield reasonable results in real-world situations. Our approach is slightly different from the standard practice of confronting each method with observed data, and then ranking the methods on the basis of various data congruence tests. As is well known, high in-sample forecasting ability does not guarantee success in making out of sample forecasts. However in making multiple forecast comparisons, the problem of data snooping has to be guarded against as pointed out recently (White (2000) and Hansen (2001)). We therefore compute Hansen's \hat{p}^{SPA} statistics with a view to assess the forecasting ability of the alternative methods that we have examined in this paper. The search for superior forecasting methods is an endless one. This article has tried to project the potentiality of a group of procedures for forecasting series with strong periodic tendencies. Of the five new methodologies presented here, Methods I, III and V seem to hold interesting promise for future applications, the first two mainly in natural

sciences applications, and the last for economic data. Needless to say, our conclusions are circumscribed by the specific data sets that we consider here, and at the moment can only be regarded as highly tentative. Only greater experience over more data sets can indicate the relative superior forecasting ability, or otherwise, of these methods *vis-à-vis* each other, as well as over other established methods.

REFERENCES

- Akaike, H. (1980), "Seasonal adjustment by Bayesian modelling," *Journal of Time Series Analysis*, 1, 1-13.
- Aparicio, F.M. (1998), "Testing linearity with information-theoretic statistics and the bootstrap," in *Nonlinear Modeling of High Frequency Financial Time Series*, eds C. Duns, and B. Zoo, Chichester, UK: Wiley
- Barnett, A.G. and Wolff, R.C. (2005), "A time domain test for some types of nonlinearity," *IEEE Transactions on Signal Processing*, 53(1), 26-33
- Bartlett, M.S. (1954), "Problèmes de L'analyse Spectrale des Series Temporelles Stationnaires," *Publication de l'Institute Statistique (University of Paris)* Fasc. Vol.3, 119-134.
- Bhansali, R.J. (1979), "A Mixed Spectrum Analysis of the Lynx Data," *Journal of the Royal Statistical Society, Ser. A*, 142(2), 199-209.
- Bhansali, R.J. and Downhan, D.Y. (1977), "Some Properties of the Order of an Autoregressive Model selected by a Generalization of Akaike's FPE Criterion," *Biometrika*, 64, 547-551.
- Bloomfield, P. (1976), *Fourier Analysis of Time Series: An Introduction* Wiley, New York.

- Brillinger, D.R. (1975), *Time Series: Data Analysis and Theory*, Holt, Rinehart and Winston, New York
- Bryson, A.E. and Ho, Y.C. (1969), *Applied Optimal Control: Optimization, Estimation and Control*, Blaisdell Pub. Co. New York
- Cadzow, J. (1982), "Spectral estimation: An overdetermined rational model equation approach," *Proceedings IEEE*, 70(9), 907-939
- Campbell, M.J. and Walker, A.M. (1977), "A Survey of Statistical Work on the Mackenzie River Series of Annual Canadian Lynx Trappings for the Years 1821-1934 and a New Analysis," *Journal of the Royal Statistical Society, Ser. A*, 140(4), 411-431.
- Candy, J.V. (1988), *Signal Processing: The Modern Approach* McGraw-Hill, New York.
- Chiu, S. (1989), "Detecting Period Components in a White Gaussian Time Series," *Journal of the Royal Statistical Society, Ser. B*, 52(2), 249-259.
- Diebold, F.X. and Mariano, R.S. (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253-263
- Durbin, J. (1960), "The Fitting of Time Series Models," *Review of the International Statistical Institute*, 28, p.233-243.
- Findley, D.F., Monsell, B.C., Bell, W.R., Otto M.C. and Chen, B. (1996), New Capabilities and Methods of the X-12 ARIMA Seasonal Adjustment Program, U.S. Bureau of the Census, (mimeo).
- Gardner, W.A. (1988), *Statistical Spectral Analysis: A Nonprobabilistic Theory*, Prentice Hall, Englewood Cliffs, NJ
- Gençay, R., Selçuz, F. and Whitcher, B. (2002), *An Introduction to Wavelets and other Filtering Methods in Finance and Economics*, London: Academic Press.
- Ghysels, E. and Osborn, D.R. (2001), *The Economic Analysis of Seasonal Time Series*, Cambridge: Cambridge University Press

- Gomez, V. and Maravall, A. (1996), "Programs TRAMO and SEATS: Instructions for the User," Bank of Spain, Working Paper No. 9628
- Grenander, U. and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series*, Wiley, New York.
- Hannan, E.J.(1971), "Non-linear Time Series Regression," *Journal of Applied Probability*, 8, 767-780.
- (1973), "The Asymptotic Theory of Linear Time Series Models," *Journal of Applied Probability*, 10, 130-145.
- Hannan, E.J., Terrell, R.D. and Tuckwell, N.E. (1970), "The seasonal adjustment of econometric time series," *International Economic Review*, 11, 24-52.
- Hansen, P.R. (2001), "An Unbiased and Powerful Test for Superior Predictive Ability," Brown University, Department of Economics, Working Paper No. 2001-06.
- Hansen, P.R. and Lunde, A. (2005), "A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?," *Journal of Applied Econometrics*, 20(7), 873-889.
- Harvey, A.C. (1989), *Forecasting Structural Time Series Models and The Kalman Filter*, Cambridge University Press, Cambridge.
- He, S. and Kedem, B. (1989), "Higher order crossings of an almost periodic random sequence in noise," *IEEE Transactions on Information Theory*, 35(3), 360-370.
- Hillmer, S.C., Bell, W.R. and Tiao, G.C. (1983), "Modelling considerations in the seasonal adjustment of economic time series," in *Applied Time Series Analysis of Economic Data* ed. A.Zellner, U.S. Department of Commerce, Bureau of the Census, Washington, DC, 74-100.

- Jakeman, A.J. and Young, P.C. (1984), "Recursive filtering and the inversion of ill-posed causal problems," *Utilitas Mathematica*, 35, 351-376.
- Kailath, T (1980): *Linear Systems*, Prentice Hall, New Jersey.
- Kalman, R.E. (1960), "A new approach to linear filtering and prediction problems," *ASME Transactions Journal of Basic Engineering*, 82-D, 35-45.
- Kay, S.M. and Marple, S.L. (1981), "Spectrum Analysis - A Modern Perspective," *Proceedings of the IEEE*, 69, 1380-1419.
- Koopmans, S.J. (1993), "Disturbance smoother for state-space models," *Biometrika*, 76, 65-79.
- Koopmans, S.J., Harvey, A.C., Doornik, J.A. and Shephard, N. (1995), *STAMP 5.0: Structural Time Series Analyzer, Modeller and Predictor*, Chapman and Hall, London.
- Kugimutzis, D. (2001), "On the reliability of the surrogate data test for nonlinearity in the analysis of noisy time series," *International Journal of Bifurcation Chaos*, 11(7), 1881-1896.
- Kundu, D. and Kannana, N. (1997), "Constrained maximum likelihood estimator for superimposed exponential signals," *Communications in Statistics - Simulation and Computation*, 26, 733-764.
- Lang, S. and McClellan, J. (1980), "Frequency estimation with maximum entropy spectral estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, 28, 716-724.
- Li, T.H. and Kedem, B. (1992), "Strong Consistency of The Contraction Mapping Method for Frequency Estimation Systems Research Centre," University of Maryland, College Park, Technical Report No. TR92-22.
- Magnus, J.R. and Neudecker, H. (1991), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley and Sons, New York.

- Marple, S. (1987), *Digital Spectral Analysis with Applications*, Prentice-Hall, Englewood Cliffs, N.J.
- Moran, P.A.P. (1953), "The Statistical Analysis of the Canadian Lynx Cycle I and II," *Australian Journal of Zoology*, 1, 163-173 and 291-298.
- Nachane, D. (2001), "Financial liberalization and monetary policy: The Indian evidence," European University Institute, Working Paper No 2001/11.
- Percival, D.B. and Walden, A.T. (2000), *Wavelet Methods For Time Series Analysis*, Cambridge (U.K.): Cambridge University Press.
- Pesaran, M.H. and Timmerman, A. (1992), "A simple nonparametric test of predictive performance," *Journal of Business and Economic Statistics*, 10(4), 461-465.
- Pisarenko, V.F. (1973), "The Retrieval of Harmonics from a Covariance Function," *Geophysical Journal of Royal Astronomical Society*, 33, 347-366.
- Politis, D.N. and Romano, J.P. (1994), "The stationary bootstrap," *Journal of the American Statistical Association*, 89, 1303-1313.
- Priestley, M.B. (1964), "Estimation of the Spectral Density Function in the Presence of Harmonic Components," *Journal of the Royal Statistical Society, Ser. B*, 26, 123-132.
- (1981), *Spectral Analysis and Time Series*, Academic Press, London.
- Prony, R. (1795), "Essai Experimental et Analytique," *Journal de L'Ecole Polytechnique*, 2, 24-76.
- Quinn, B.G. and Thomson, P.J. (1991), "Estimating the Frequency of a Periodic Function," *Biometrika*, 78(1), 65-74.
- Rice, J.A. and Rosenblatt (1988): "On frequency estimation", *Biometrika*, 75(3), 477-484.

- Sakai, H. (1984), "Statistical analysis of Pisarenko's method for sinusoidal frequency estimation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, 32, 95-101.
- Schuster, A. (1898), "On the Investigation of Hidden Periodicities with Applications to a Supposed 26 day Period of Meteorological Phenomena," *Terrestrial Magnetism*, 3, 13-41.
- Slutsky, E. (1937), "The Summation of Random Causes as the Source of Cyclic Processes," *Econometrica*, 5, 105-146.
- Smyth, G.K. (2000), "Employing symmetry constraints for improved frequency estimation by eigenanalysis methods," *Technometrics*, 42(3), 277-289.
- Stewart, G.M. (1973), *Introduction to Matrix Computations*, Academic Press, New York.
- Stoica, P. and Nehorai, A. (1988), "Study of the statistical performance of the Pisarenko harmonic decomposition method," *Communications, Radar and Signal Processing*, 153, 161-168.
- Sullivan, R., Timmerman, A. and White, H. (1999), "Data-snooping, technical trading rules and the bootstrap," *Journal of Finance*, 54, 1647-1692.
- Truong-Van, B. (1990), "A New Approach to Frequency Analysis with Amplified Harmonics," *Journal of the Royal Statistical Society, Ser. B*, 52(1), 203-221.
- Waldheimer, M. (1961), *The Sunspot Activity in the Years 1610-1960* Schulthess, Zurich.
- Walker, A.M. (1973), "On the Estimation of a Harmonic Component in a Time Series with Stationary Dependent Residuals," *Advances in Applied Probability*, 5, 217-241.

- West, K.D. (1996), "Asymptotic inference about predictive ability," *Econometrica*, 64, 1067-1084.
- White, H. (2000), "A reality check for data snooping," , *Econometrica*, 68, 1097-1126.
- Whittle, P. (1952), "Tests of Fit in Time Series," *Biometrika*, 39, 309-318.
- Yakowitz, S. (1991), "Some contributions to a frequency location method due to He and Kedem," *IEEE Transactions on Information Theory*, 37, 1177-1181.
- Young, P.C. (1991), "Comments on likelihood and cost as path integrals," *Journal of the Royal Statistical Society, Ser. B*, 53, 529-531.
- Young, P.C. and Pedregal, D.J. (1999), "Recursive and en-bloc approaches to signal extraction," *Journal of Applied Statistics*, 26, 103-128.
- Young, P.C., Pedregal, D.J. and W.Tych (1999), "Dynamic harmonic regression," *Journal of Forecasting*, 18, 369-394.
- Yule G.U. (1927), "On a Method of Investigating Periodicities in Disturbed Series with Special reference to Wolfer's Sunspot Numbers," *Philosophical Transactions of the Royal Society, Ser.A*, 226, 267-298.

APPENDIX: TABLES

Table A. 1: Harmonic Frequencies for Sunspot Data.

	Method I	Method II	Method III	Method IV	Method V
	0.0687	0.0736	0.1195	0.1083	0.0863
	0.1164	0.3682	0.5375	0.2874	0.1079
Frequencies	0.2258	0.4172	0.6280	0.4561	0.2840
	0.5653	0.5645	3.1082	0.8926	0.3759
	0.6251	2.2089	—	1.4571	0.6819
	—	2.3317	—	2.0153	

Note:

In Methods II and IV, the number of harmonics is estimated via Chiu's method

Table A. 2: Forecasting Comparisons using Hansen's Superior Predictive Ability (SPA) Test for Sunspot Data

Benchmark	$\hat{p}^{SPA}(i)$	$\hat{p}^{SPA}(i)$	$\hat{p}^{SPA}(i)$	$\hat{p}^{SPA}(i)$
Method	RMSE	MAPE	MAXAE	LS
Method I	0.7219	0.6632	0.6203	0.4821
Method II	0.0738	0.05380	0.0019**	0.0669
Method III	0.6854	0.7302	0.8138	0.3408
Method IV	0.1348	0.2146	0.1453	0.3778
Method V	0.5768	0.4193	0.4704	0.3576
Method VI	0.0473*	0.0135*	0.0918	0.0082**

Notes:

- (i) For definitions of various methods see Section 1. Method VI is the random walk model.
- (ii) (*) and (**) denote significance at 5% and 1% respectively
- (iii) $\hat{p}^{SPA}(i)$; $i = I, II, \dots, VI$ denote the values of Hansen's test statistic for the SPA test, when the i -th method is considered as the benchmark method.

Table A. 3: Harmonic frequencies for Spanish IIP Data

	Method I	Methods II	Method III	Methods IV	Method V
Frequencies	0.0843	0.1852	0.0533	0.2359	0.1211
	0.2849	0.3435	0.3946	0.4462	0.2653
	0.4712	0.5106	0.8514	0.5352	0.4366
	0.6538	0.7680	—	0.9111	1.5714
	1.0288	1.9059	—	1.4803	—
	2.51	2.6837	—	2.1562	—

Note:

In Methods II and IV, the number of harmonics is estimated via Chiu's method

Table A. 4: Forecasting Comparisons using Hansen's Superior Predictive Ability (SPA) Test for Spanish IIP Data

Benchmark	$\hat{p}^{SPA}(i)$	$\hat{p}^{SPA}(i)$	$\hat{p}^{SPA}(i)$	$\hat{p}^{SPA}(i)$
Method	RMSE	MAPE	MAXAE	LS
Method I	0.6181	0.5605	0.4922	0.3218
Method II	0.0613	0.1945	0.0026**	0.0436*
Method III	0.4710	0.5394	0.7026	0.3153
Method IV	0.2493	0.2801	0.1857	0.4126
Method V	0.8026	0.8492	0.6471	0.4936
Method VI	0.0460*	0.0389*	0.0826	0.0067**

Notes:

- (i) For definitions of various methods see Section 1. Method VI is the random walk model.
- (ii) (*) and (**) denote significance at 5% and 1% respectively
- (iii) $\hat{p}^{SPA}(i)$; $i = I, II, \dots, VI$ denote the values of Hansen's test statistic for the SPA test, when the i -th method is considered as the benchmark method.